2

RADC-TR-90-404, Vol XVII (of 18)
Final Technical Report
December 1990

HIGH-LEVEL ADAPTIVE SIGNAL PROCESSING

Northeast Artificial Intelligence Consortium (NAIC)

Hamid Nawab and Victor Lesser

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

This effort was funded partially by the Laboratory Director's fund.

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the RADC Public Affairs Division (PA) and is releasable to the National Technical Information Services (NTIS) At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-90-404, Volume XVII (of 18) has been reviewed and is approved for publication.

APPROVED: Vincent Vannicola

VINCENT VANNICOLA
Project Engineer

APPROVED:

JAMES W. YOUNGBERG, Lt Col, USAF
Deputy Director of Surveillance

FOR THE COMMANDER: Igor G. Plonisch

IGOR G. PLONISCH
Directorate of Plans & Programs

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (OCTS) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE December 1990	3. REPORT TYPE AND DATES COVERED Final Sep 84 - Dec 89

4. TITLE AND SUBTITLE
HIGH-LEVEL ADAPTIVE SIGNAL PROCESSING

5. FUNDING NUMBERS
C - F30602-85-C-0008
PE - 62702F
PR - 5581
TA - 27
WU - 13
(See reverse)

6. AUTHOR(S)
Hamid Nawib and Victor Lesser

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Northeast Artificial Intelligence Consortium (NAIC)
Science & Technology Center, Rm 2-296
111 College Place, Syracuse University
Syracuse NY 13244-4100

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Air Development Center (COES)
Griffiss AFB NY 13441-5700

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
RADC-TR-90-404, Vol XVII
(of 18)
(See reverse)

11. SUPPLEMENTARY NOTES
RADC Project Engineer: Vincent Vannicola/OCTS/(315) 330-4437

This effort was funded partially by the Laboratory Director's fund.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The Northeast Artificial Intelligence Consortium (NAIC) was created by the Air Force Systems Command, Rome Air Development Center, and the Office of Scientific Research. Its purpose was to conduct pertinent research in artificial intelligence and to perform activities ancillary to this research. This report describes progress during the existence of the NAIC on the technical research tasks undertaken at the member universities. The topics covered in general are: versatile expert system for equipment maintenance, distributed AI for communications system control, automatic photointerpretation, time-oriented problem solving, speech understanding systems, knowledge base maintenance, hardware architectures for very large systems, knowledge-based reasoning and planning, and a knowledge acquisition, assistance, and explanation system.

The specific topic for this volume is the integration of Artificial Intelligence and Signal Processing for allocating processing resources and adjusting parameters of low-level processing in accordance with evolving high-level interpretations of the signal-generating environment.

14. SUBJECT TERMS
Artificial Intelligence, Signal Processing, H-LASP, High-Level Adaptive Signal Processing

15. NUMBER OF PAGES
52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Block 5 (Cont'd)   **Funding Numbers**

PE - 62702F	PE - 61102F	PE - 61102F	PE - 33126F	PE - 61101F
PR - 5581	PR - 2304	PR - 2304	PR - 2155	PR - LDFP
TA - 27	TA - J5	TA - J5	TA - 02	TA - 27
WU - 23	WU - 01	WU - 15	WU - 10	WU - 01

Block 11 (Cont'd)

This effort was performed as a subcontract by the University of Massachusetts at Amherst to Syracuse University, Office of Sponsored Programs.

VOLUME 17

High-Level Adaptive Signal Processing – Final Report to RADC

Hamid Nawab Victor Lesser

November 20, 1990

Contents

1 Executive Summary

In this report, we describe the accomplishments of the high-level adaptive signal processing (H-LASP) project, carried out by a team of researchers from the University of Massachusetts at Amherst and Boston University during the period from February 1989 to September 1989. High-level adaptive signal processing (H-LASP) involves the integration of artificial intelligence and signal processing in an interpretation system and makes use of a paradigm that allocates processing resources and adjusts parameters of the low-level processing in accordance with the evolving high-level interpretations of the signal-generating environment. The goal of the project reported here was to evaluate how the H-LASP paradigm applies to a realistic task: real-time sound classification. We have built a testbed for this application and found that with some modifications and a number of refinements, the H-LASP paradigm can be successfully used for the development of signal interpretation systems.

In high-level adaptive signal processing, the integration of high and low level processing is achieved through a problem-solving paradigm that involves three phases: discrepancy detection, diagnosis, and signal re-processing through control parameter adjustment. Discrepancy detection is carried out by comparing the features of the signal processing outputs with features expected on the basis of the evolving scenario interpretation and with a-priori knowledge about the signal-generating environment. This is followed by a diagnostic reasoning process that makes significant use of the underlying Fourier theory of the signal processing system to isolate a subset of system parameters whose settings were likely to have caused the observed discrepancies. Finally, the signal processing resources are reallocated by appropriately adjusting system parameters in order to re-process the input signal with the aim of removing the observed discrepancies. This paradigm was established in our previous research on an acoustic localization problem where we had found that expert human signal processors use this type of reasoning in manually reallocating the signal processing resources through parameter adjustment. The need for resource allocation for the low-level processing components arises because of two factors. The *model variety* factor is that the signal processing resources (which are always finite) have to deal with an infinite variety of signal classes whose signal processing requirements are often in conflict with each other. By adjusting the parameters of an algorithm it can be made to deal with different classes of signals. The second factor that leads to the need for resource allocation is the *real-time performance* factor. In a real-time situation, there is not always enough time to do all the signal processing the system would ideally carry out. In such cases, focus-of-attention decisions have to be made about the use of the signal processing resources within the available time frame.

For the project described in this report, the goal was to evaluate and improve the H-LASP paradigm for a practical sound classification application. We selected the real-time sound classification problem for this purpose because it offers two major advantages: (1) it shares many low-level and high-level processing requirements with other signal interpretation problems such as radar signal interpretation and (2) the acoustic signal database is readily available in our university laboratories for testbed experiments. The specific sound classification problem arises in the context of real-time interpretation of acoustic signals received by a system (robot, if you will) stationed in a household environment. This means that the various sounds being received by the system have to be classified in terms of the sources from which these sounds originate. In the household environment, we are interested in sources such as telephones, vacuum cleaners, babies, speech, footsteps, doorbells etc. The problem is made particularly complicated (thereby requiring Artificial Intelligence techniques at the higher levels) because several sources may occur simultaneously and they may have overlapping frequency spectra.

The achievements of our project may be divided into five major categories:

- Incorporation of the diagnostic reasoning process into the sound classification testbed along with refinements in that process to deal with the more sophisticated theory underlying the new application.

- Formulation and implementation of a practical approach to discrepancy detection for the sound classification task.

- Implementation in the testbed of a sophisticated database using the Generic Blackboard (GBB) system. The design of the database within a blackboard framework was found to ease the development of the processing components of the H-LASP paradigm in the form of independent knowledge sources.

- Design of the control component of the testbed through adaptation of a framework developed at the University of Massachusetts for the control of interpretation through analysis of the sources of uncertainty associated with the various evidence gathering mechanisms.

- Design of the control component of the testbed to ensure real-time invocation of the high and low-level knowledge sources while maintaining the integrity of the high level interpretations to within the goals of the system.

Within its limited eight-month duration, the project was successful in developing a testbed that includes a blackboard database with knowledge sources for signal processing, signalre-processing, discrepancy detection, and diagnosis. Although the parameter adjustment and system control components were fully designed, further work is needed to complete the implementation of the parameter adjustment knowledge sources and the control component of the system. Completion of these components will permit us to thoroughly evaluate the performance of a fully integrated H-LASP system for a practical real-time signal interpretation application.

2 Ancillary Activities

2.1 Publications

1). I. Gallestegui et. al. Implementing a Blackboard-based Sound Classification System: A Case Study. Proceedings of the Blackboard Workshop at IJCAI 89. Detroit, MI. August 1989.

2). F. Klassner et. al. A Computer Program for the Symbolic Processing of Sound Spectra. Submitted to the 1990 International Conference on Acoustics, Speech, and Signal Processing.

2.2 Presentations

Hamid Nawab. *High-Level Adaptive Signal Processing.* Biomedical Engineering Graduate Seminar. Boston University. April 1989.

Victor Lesser. *High-Level Adaptive Signal Processing.* Fifth Annual Workshop of the AI Consortium. August 1989.

3 Introduction

The *long-term* goal of our research is the establishment of a systematic framework for the *integration* of artificial intelligence concepts and techniques into complex signal processing systems in order to make their behavior more adaptive to the high-level characteristics of the signal-generating environment. This is in contrast to most present-day complex signal processing systems, where if there is any artificial intelligence, it usually comes *after* the signal processing has been completed [1,2]. Whereas signal processing is most often a real-time activity, the interpretation of the outputs of such signal processing is either over-simplified because of real-time constraints or it is not carried out in real-time. In either case, it has been considered unrealistic for the higher-level processing to affect the way the real-time signal processing is carried out. However, continuing advances in hardware and artificial intelligence technology have now made it practical to consider the design of systems in which the higher-level processing is sophisticated enough and fast enough to influence the real-time use of signal processing resources.

The goal of the H-LASP project was to refine the H-LASP paradigm of discrepancy detection, diagnosis, and signal re-processing through parameter adjustment in the context of a real-time signal processing and interpretation application. The acoustic localization research had focused on the nature of the reasoning performed by experts while they were determining how to adjust the signal processing parameters, but that research had not considered the problem of how such a system would form expectations about what is likely to happen in the signal-generating environment so that this information may be used for discrepancy detection when compared to the actual signal processing outputs. In the H-LASP project, we have studied the problem of discrepancy detection and formulated a variety of solutions, described in this report. Another goal of the H-LASP project was to test the applicability of the diagnostic reasoning process that we had formulated in the acoustic localization research to the sound classification problem. We were successful in incorporating the diagnostic reasoning process into the sound classification testbed and we were able to make further refinements in how the process deals with the more sophisticated signal processing theory underlying the new application. Further details are included in the section on diagnostic reasoning in this report. A third objective of our project was to further refine the qualitative reasoning aspects of the H-LASP paradigm. Because the system has to deal with various amounts of uncertainties and error in the data it handles, it is necessary to reason with qualitative specifications of many of the quantities. In particular, during this project we came to the conclusion that an important enhancement to the H-LASP paradigm is to include a control strategy framework that controls the system's resources in accordance with the importance of the uncertainties in the interpreted data. For this purpose, we adopted a framework [3] developed at the University of Massachusetts for the control of interpretation through analysis of the sources of uncertainty associated with the various evidence gathering mechanisms. As the H-LASP paradigm evolved into a more complex framework, we also found the need for more attention to be given to the representation of data and knowledge contained in the system. We opted for a blackboard framework which has the advantage of dividing the database into as many levels of abstraction as needed and to separate the development of knowledge sources in accordance with the levels at which they operated. In the testbed, we used the Generic Blackboard System (GBB) [4] as the shell for developing the specific application blackboard. Details of the blackboard architecture and the implementation issues we faced during the development are included in the report. In integrating knowledge sources with the blackboard, we also had to incorporate into the overall system design considerations arising from the real-time nature of the sound classification application. Although our testbed cannot operate in real-time because of hardware limitations, the design of the processing activity is such that with appropriate hardware, the system can operate in real-time. A discussion of the considerations for real-time processing is included in the report.

The remainder of the report is organized as follows. In section 4, we give the background of how previous work on acoustic localization led to the formulation of the H-LASP paradigm. The sound classification problem in the context of which our H-LASP testbed was developed is described in section 5. This is followed in section 6 with a description of the signal processing resources utilized in our sound classification testbed. In sections 7-11, we provide details of the various issues encountered during the project regarding the design of the blackboard database, discrepancy detection, diagnosis, resource allocation and parameter adjustment, and real-time operation.

4 Background

Prior to the project described in this report, our own work in the area of acoustic localization [5] indicated the importance of tighter integration between artificial intelligence and signal processing. We concentrated on signal processing systems that have an underlying mathematical theory, largely in the Fourier frequency domain. Such systems often have a large number of parameters that need to be adjusted in accordance with certain high-level characteristics of the signal-generating environment. In our acoustic localization application, the signal-generating environment consisted of aircraft flyby's, recorded on acoustic microphones. Typically, such systems have their parameter settings fixed for the "average scenario." Since the acoustic characteristics of various aircraft differ from each other and since the number of aircraft present (and their relative locations) within the range of the microphones is highly variable, the fixed parameter settings are not appropriate in all situations. For example, when two aircraft are within the range of the microphones, whether or not the signals can be used to localize and classify each of the aircraft depends to a large extent on the temporal and spatial frequency *spectra* of the signals generated by the two aircraft. It is often the case that the temporal spectra of the two aircraft overlap to a certain extent. Therefore it is necessary for the signal processing system to focus on the non-overlapping frequency regions in order to differentiate between the two aircraft. The spatial frequency information received at the microphones is highly dependent on the relative locations of the two aircraft at each instant. In certain situations, it becomes difficult to distinguish the directionality of the received signals unless the signal processing has a-priori knowledge or an expectation about the temporal frequency characteristics of the individual aircraft. This a-priori knowledge may then be used to tailor the spatial processing for the purpose of extracting directional information. We see, therefore, the importance of *controlling* the parameters of the signal processing system in response to a higher-level interpretation or expectation of the signal-generating environment.

In light of our experience with the acoustic localization problem, the project described in this report was formulated with the aim of further developing the concept of high-level adaptive signal processing on the basis of a paradigm whose major components are *Discrepancy-Detection, Diagnosis, and Signal-Reprocessing with Parameter Adjustment.* For the acoustic localization problem, we had found that human experts adjusted the signal processing resources by searching for discrepancies between features of the actual signal processing outputs and features expected on the basis of a-priori knowledge about the signal generating environment (we refer to this as discrepancy detection). This was followed by a reasoning process that made significant use of the underlying Fourier theory of the signal processing system to isolate a subset of system parameters whose settings were likely to have caused the observed discrepancies (this constitutes the diagnosis part of the paradigm). Finally, the isolated parameters are adjusted with the aim of removing the observed discrepancies (this is the parameter adjustment part of the H-LASP paradigm).

The acoustic localization project helped us to formulate the discrepancy detection, diagnosis, and parameter-adjustment mechanisms for signal re-processing as the basis of a high-level adaptive signal processing system design. To demonstrate the concepts involved in our system design, in the next section we present an example to illustrate how such a system operates in a particular situation.

5 Demonstration of Concept

In this section, we present an example of how the H-LASP paradigm is used to carry out the processing required for the interpretation of a signal that is a linear combination of signals from different sources with different characteristics in time and frequency. This type of situation often arises in the context of signal classification problems. The details of the H-LASP sound classification testbed that carries out such processing are given in later sections.

Let us consider a twelve-second acoustic signal S that we wish to process in order to determine the time-varying frequency content of the component signals that are due to different sources. In figure 1, we show the actual time-frequency characteristics of the signal S. The signal contains contributions due to four sources, S1, S2, S3, and S4. Source S1 is a low-frequency monochromatic signal that lasts for the entire duration of the 12-second signal S. Source S2 gives rise to a frequency-modulated monochromatic signal that lasts approximately from the first second to the ninth second. Source S3 contains two harmonics lasting from approximately the sixth second to the twelfth second. Note that the two components of S3 have an abrupt change in frequency during the ninth second. Source S4 contains five harmonics which begin shortly after the ninth second and last for approximately two seconds. In our testbed, the signal data to be processed arrives in two-second intervals demarcated by the dashed vertical lines in figure 1.

When the first two-second frame of signal-data undergoes front-end short-time Fourier transform (STFT) signal processing [1]to determine its time-dependent frequency content, the result obtained is shown in figure 2. In particular, note that while the frequency content due to S1 is captured, there is no contribution due to S2. The testbed front-end signal processing also consists of time-domain (TD) processing to measure the energy and the zero crossing rate in the waveform. The results of the TD processing are used to check for consistency with the STFT results. In the case of the results for the first frame, the testbed finds a significant difference in the energy measurement from the TD process and the energy in the STFT output. This

[1] STFT processing includes peak detection which uses an energy threshold to reject peaks whose energies are lower than the threshold.

type of discrepancy is referred to as a data-data discrepancy since it results from comparing the output data of two different signal processing algorithms applied to the same underlying signal. The existence of this discrepancy triggers a *Diagnosis* knowledge-source in the testbed. This knowledge source is used to hypothesize the cause for the discrepancy. In this situation, the Diagnosis knowledge source that we have designed correctly hypothesizes that the energy discrepancy is due to the fact that the energy threshold used for detecting peak tracks in the STFT was too high. Conequently, the system decides to decrease the threshold by a factor of 1/2 and re-process the signal in the first frame.

The result of the first signal re-processing on the first frame is shown in figure 3. In this case, we observe that although the frequency track due to source S2 has been detected, there are some additional short tracks in the STFT output. The higher level interpretation knowledge sources attempt to find a consistent explanation for those short tracks and fail to find any such explanations. This is referred to as a data-interpretation discrepancy. The *Diagnosis* knowledge source is triggered. It determines that the short "noise" tracks may be eliminted by raising the peak detection threshold in the STFT processing in such a way that only the two highest energy tracks are detected. The consequent second round of signal re-processing results in the output shown in Figure 4. The higher level interpretation knowledge sources are able to classify the frequency track S1 as being due to a specific target type A. On the other hand, the track due to S2 is classified as belonging to a class of targets rather than a specific target. This is because the observed track for S2 is determined to potentially belong to a variety of different target types. To remove some of the other possibilities, a search is conducted for specific frequency tracks that would have to be present in the first frame along with the observed track. These might be low energy tracks, but energy thresholding is not needed in this case because the frequency tracks are searched for in the specific frequency regions as dictated by the corresponding target models. The third round of signal re-processin thus involves a search for specific frequency tracks in frame 1. However, no such frequency tracks are found, as indicated in Figure 5. Now the remaining uncertainty about the identity of the target corresponding to S2 can be resolved only by waiting for more waveform data to arrive.

Since it is essential to continue tracking the frequency content due to S2 in the second frame, the *Global parameter Adjustment* knowledge source in the testbed decides to use for the front-end signal processing of the second frame the signal processing control parameter values that were used in the second round of signal re-processing of the first frame. The results, illustrated in figure 6, for the front-end signal processing in the second frame are found to be sufficient to uniquely classify S2 as belonging to a target of a a specific type B. The model for that target as stored in the system's knowledge base, indicates that the target has a periodic frequency modulation. The system thus forms an "expectation" for the future evolution of S2. These expectations are matched when the results from the front-end processing of the third frame are obtained (see Figure 7).

The result after the front-end signal processing of the fourth frame is shown in Figure 8. A new track is obtained in the lower frequency region. TD analysis of the waveform in that frequency region (obtained through bandpass filtering) reveals that the zero-crossing rate is not compatible with a monochromatic source. This data-data discrepancy results in the application of the *Diagnosis* knowledge source, which hypothesizes that there is a frequency-resolution problem in that frequency band. The signal re-processing planner responds by suggesting that the frequency resolution of the STFT be increased by increasing the value of the STFT window-length control parameter and decreasing the peak detection energy threshold. The consequent signal re-processing result for the fourth frame is shown in Figure 9. Note that now the two tracks due to S3 (see Figure 1) have been resolved. On the other hand, part of the S2 track is missing because of the decreased time resolution when the STFT window length is increased. However, the system uses the results of the fronnt-end signal processing of the fourth frame to conclude that S2 is still present. Also the interpretation knowledge sources associate S3 with a target class C, with uncertainty due to the fact that the entire temporal data for S3 has not yet been receieved.

The result of front-end signal processing for frame 5 is shown in Figure 10. Once again a data-data discrepancy indicates a frequency resolution problem. After signal re-processing, the result is shown in figure 11. Note that the extra harmonics of S4 are now detected, although time-resolution problems cause the frequency modulated track of S3 to be missed. However, that information is available to the system from the results of front-end signal processing for frame 5. There is now enough information to classify S3. However, there is not enough information to classify S4. That uncertainty is resolved after the front-end signal processing in frame 6. At that point the entire 12 second signal has been successfully interpreted.

The above example illustrates the kind of interpretation that takes place in the H-LASP testbed in the context of sound understanding. The next section presents some background on the sound understanding testbed. It is followed by sections that describe various architectural aspects of the H-LASP testbed.

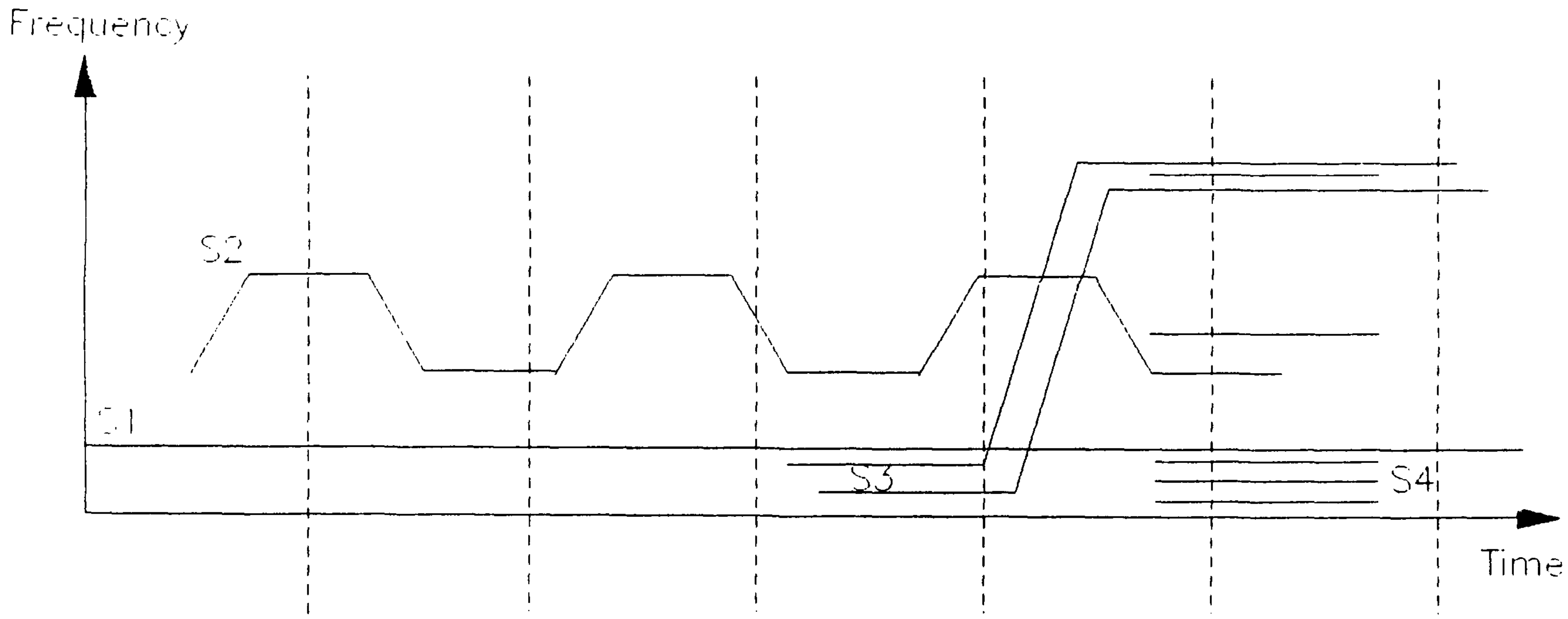
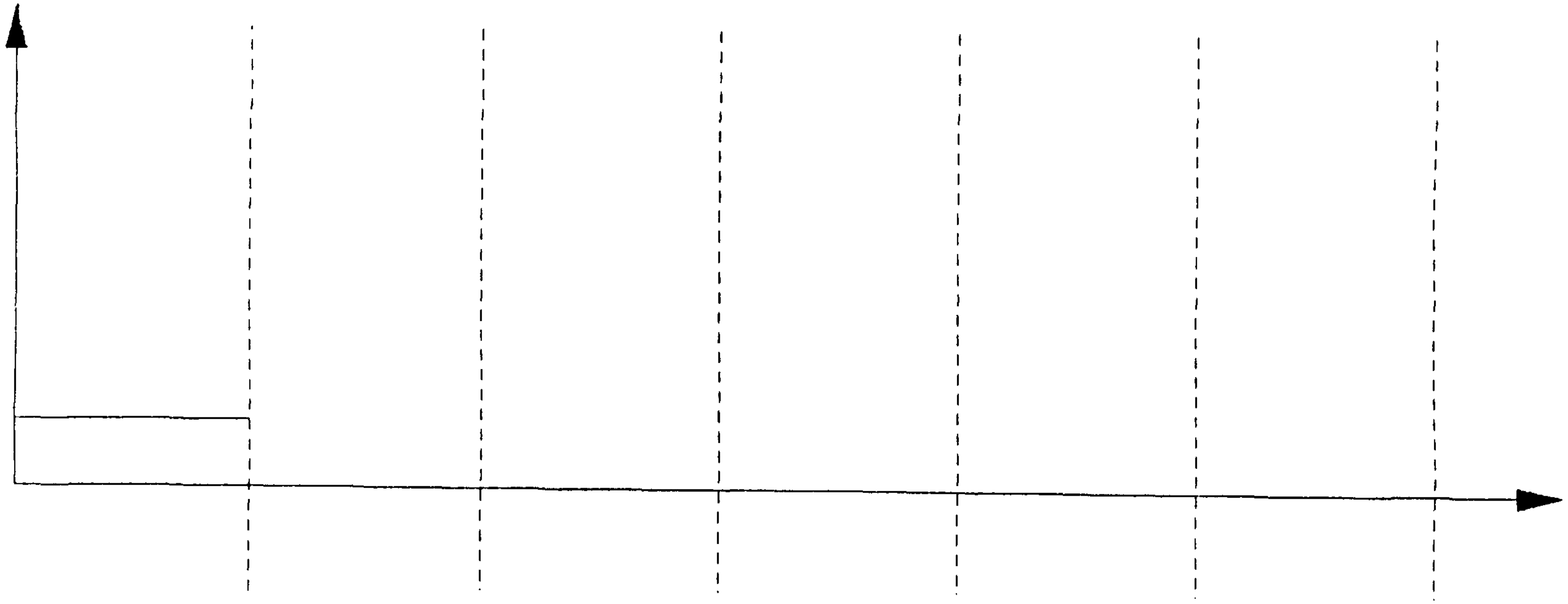
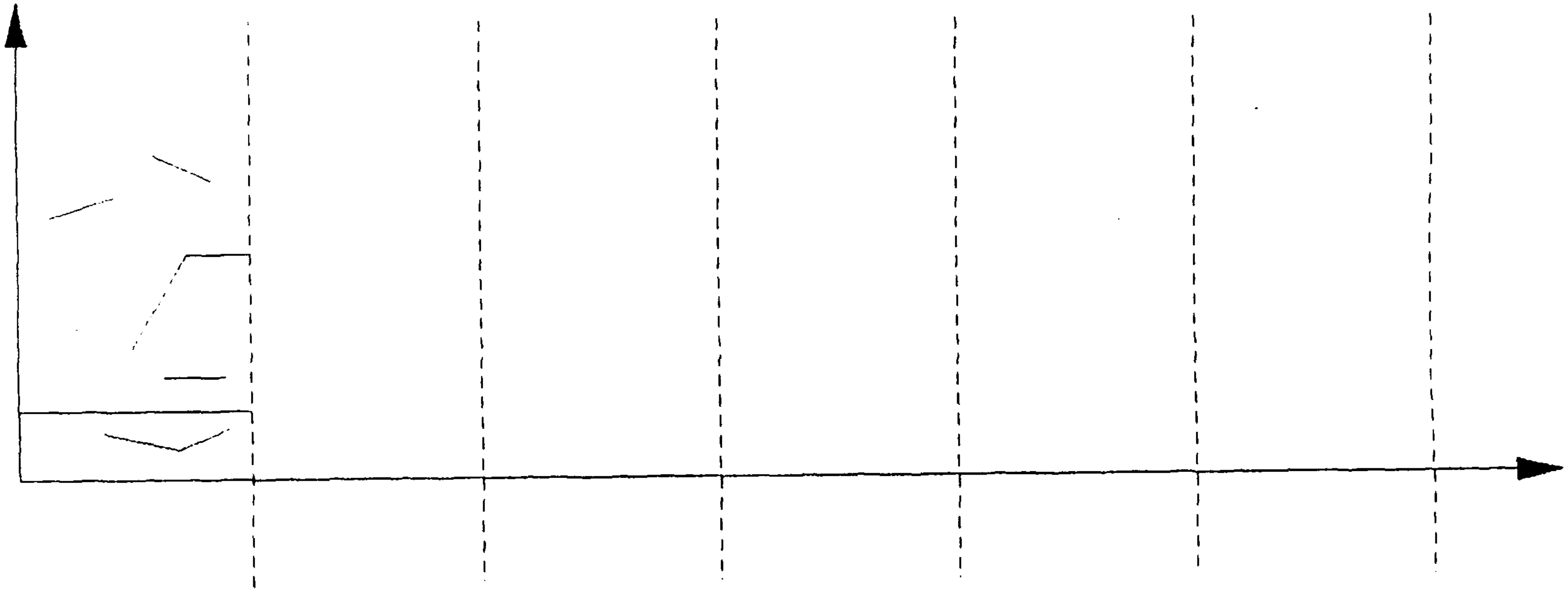


Figure 1: Actual Scenario.



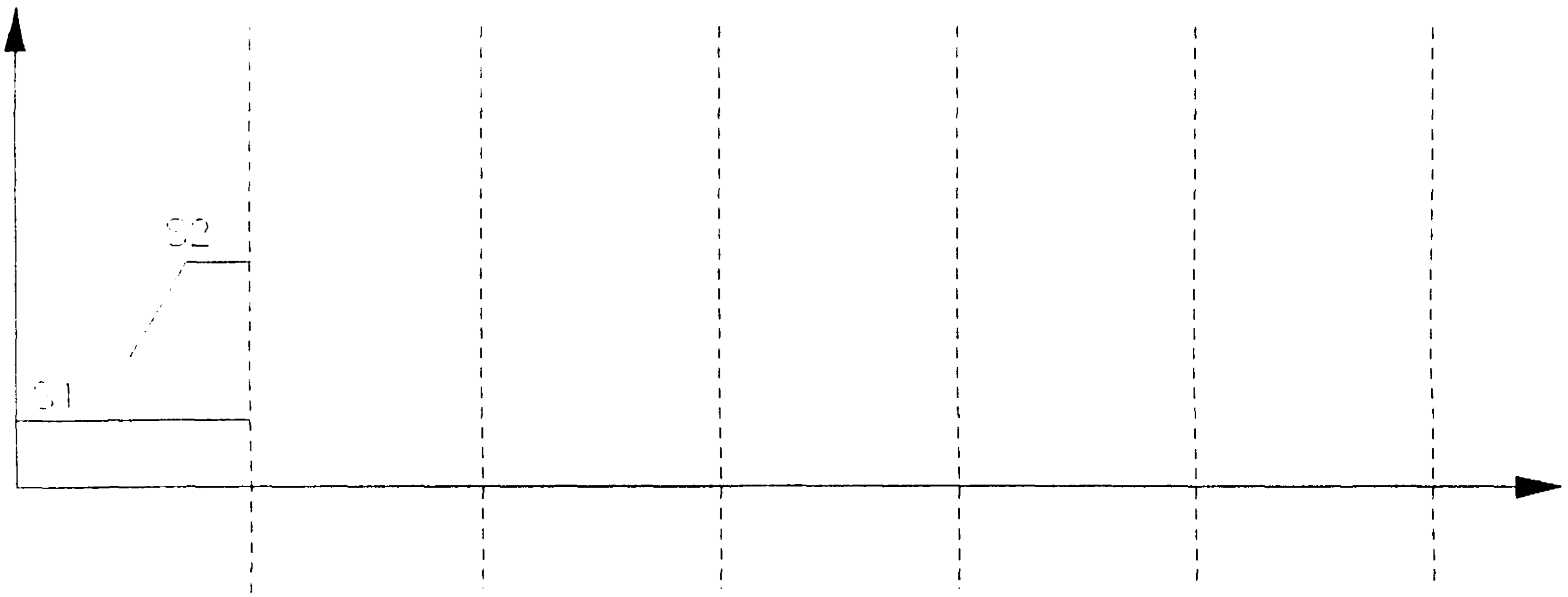
Data – Data Discrepancy between energy in T.D. and energy in peak track
Diagnosis: STFT peak threshold too high
Signal Re–processing with threshold halved

Figure 2: Front-End Signal Processing in Frame One.



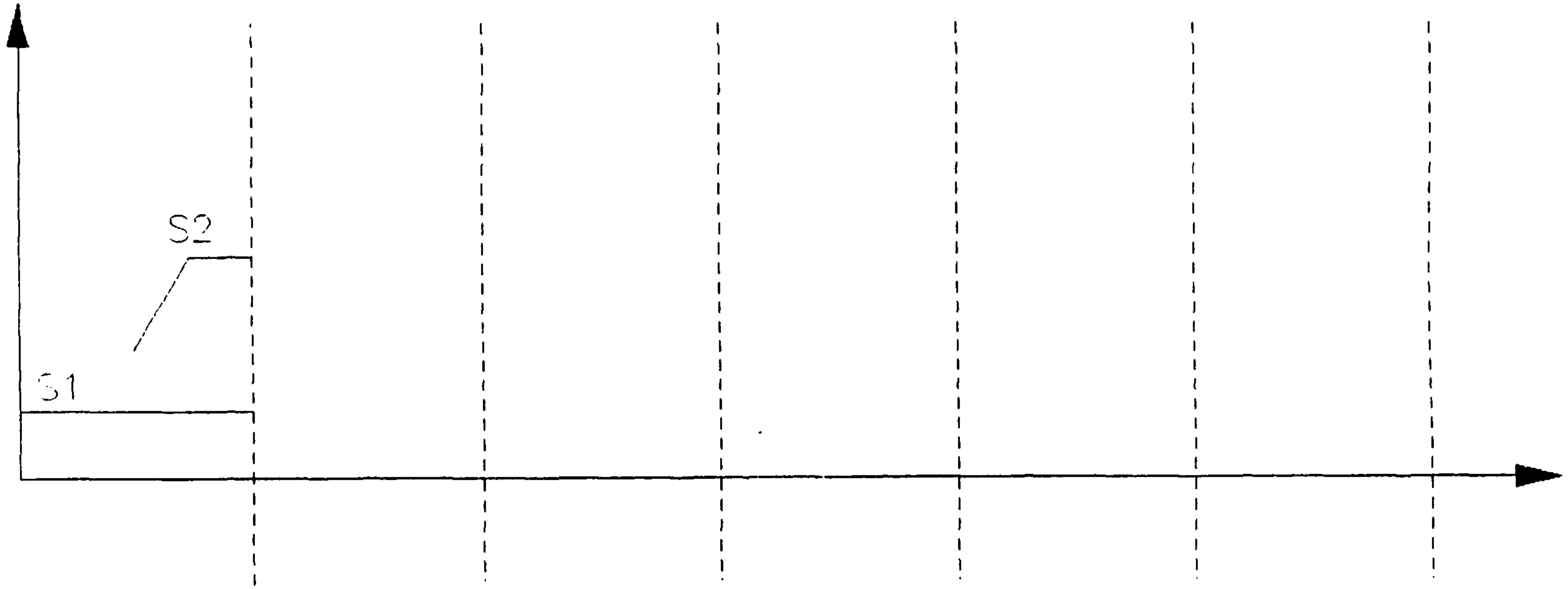
Data – Interpretation Discrepancies. Noise tracks.
Diagnosis: Too many peaks being selected.
Signal re–processing with two highest energy peaks.

Figure 3: Frame One After the First Signal Re-processing.



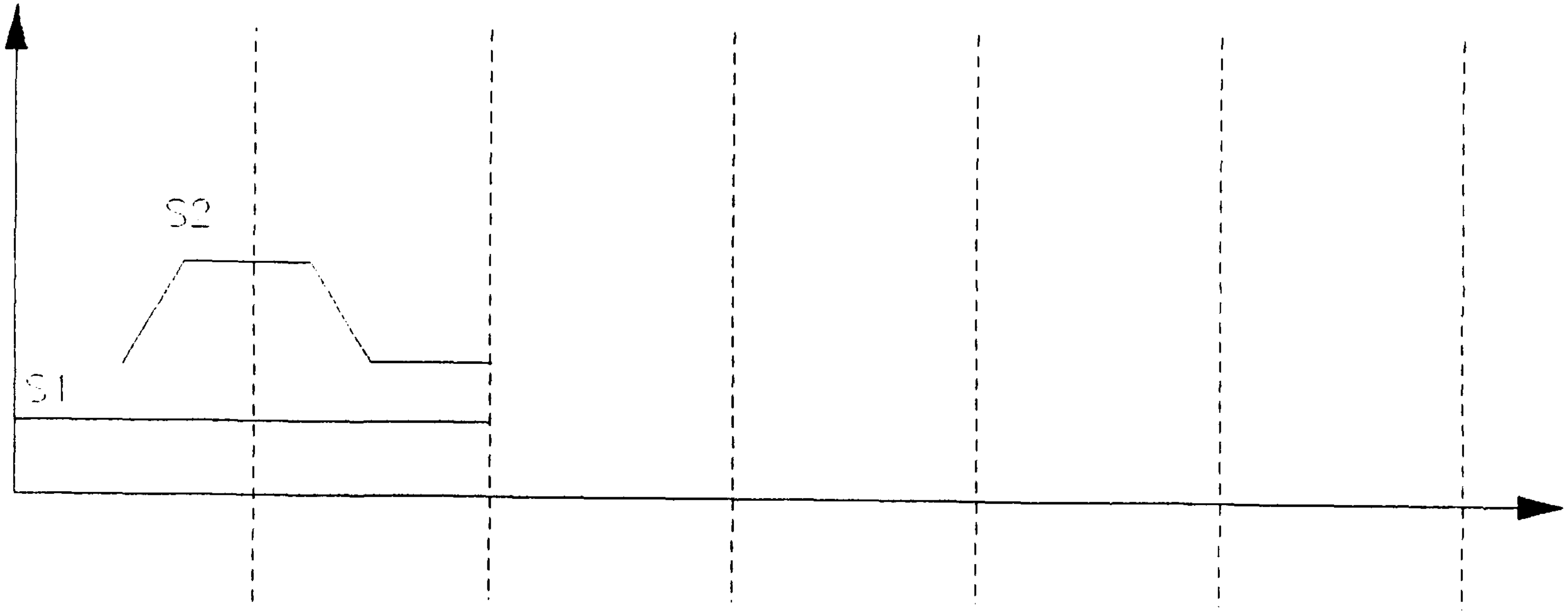
S1: classified as Target A with no uncertainty.
S2: classified as Target Class B. Uncertainties are due to
incomplete temporal data, partial frequency data.
Signal re-processing to search for specific frequencies.
to prune class B hypotheses for S2.

Figure 4: Frame One After the Second Signal Re-processing.



Multiple frequency elements of Class B pruned out.
To remove uncertainty due to incomplete temporal data proceed to next frame.

Figure 5: Frame One After the Third Signal Re-processing.



Global Parameter Adaptation to continue tracking S2.
S2: classified as Target B because sufficient temporal data.
S2: expected to periodically replicate.

Figure 6: Front-End Signal Processing in Frame Two.

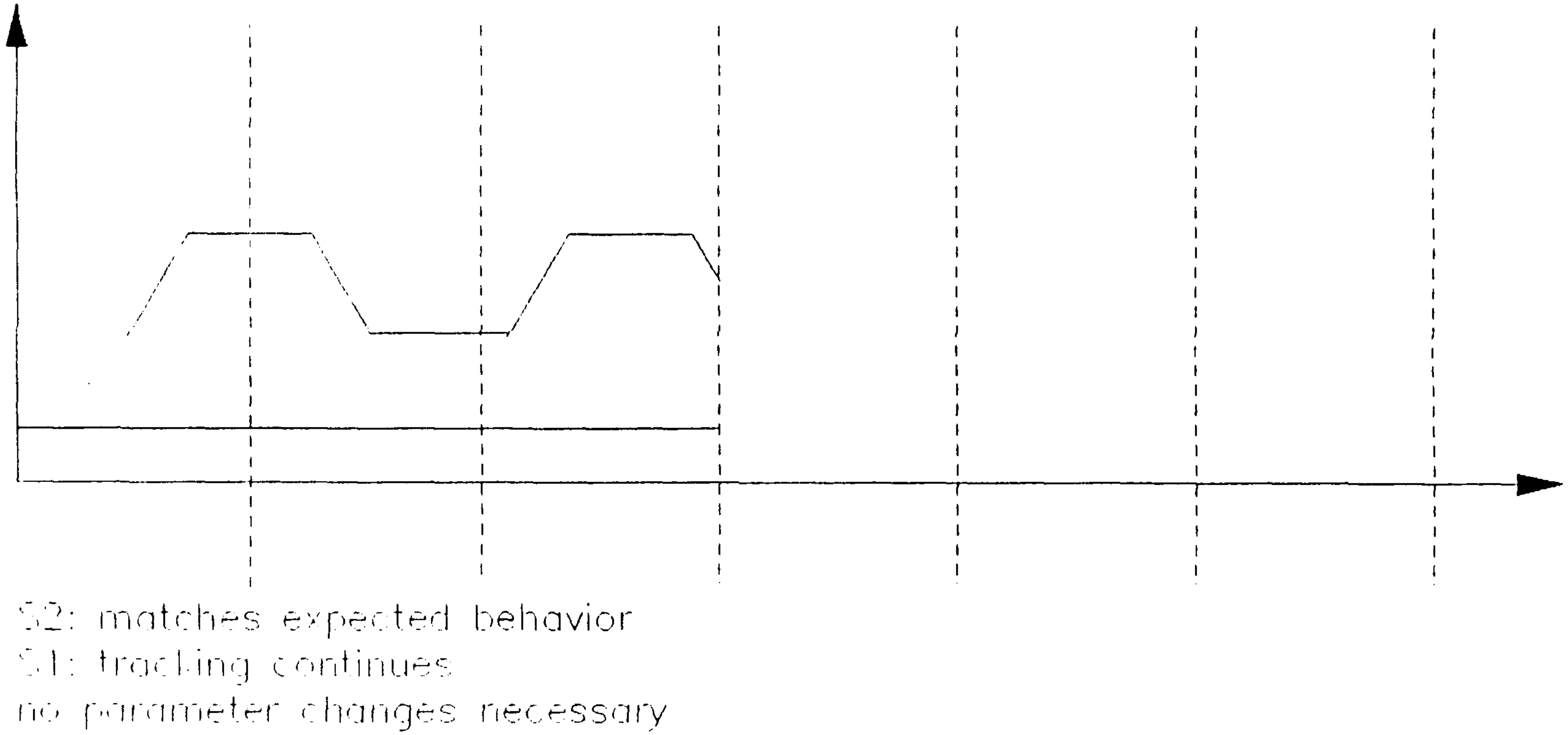
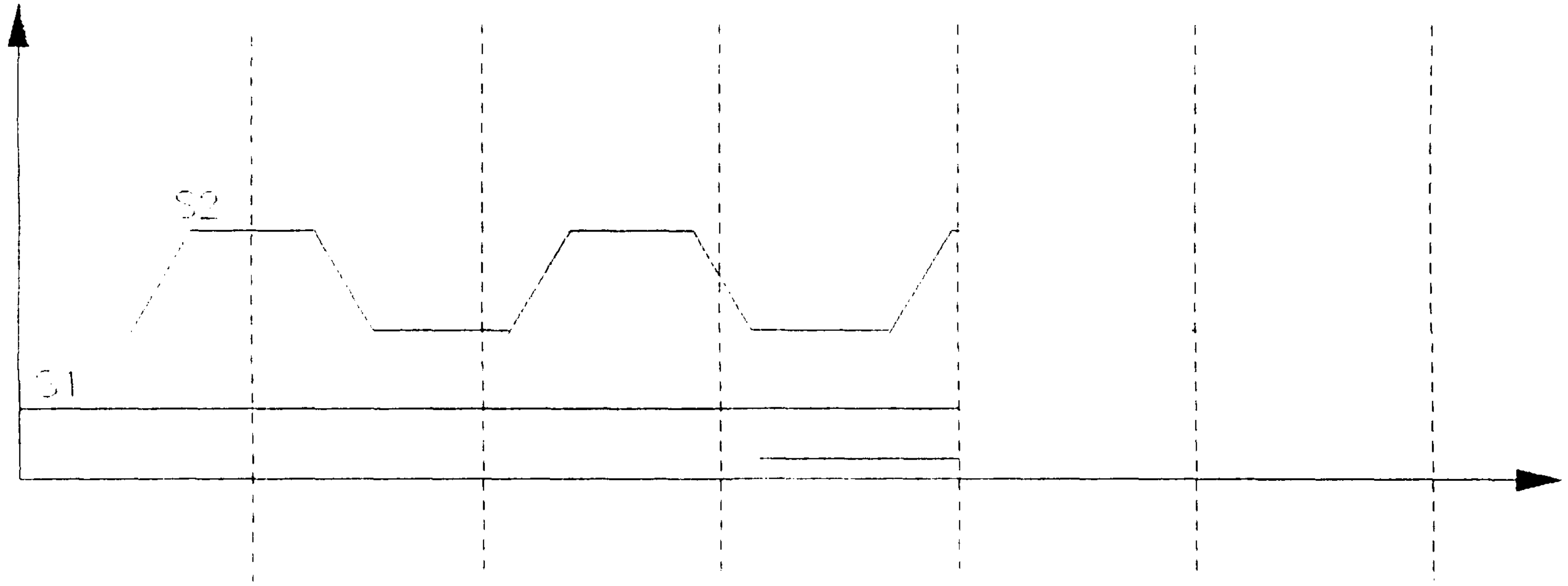
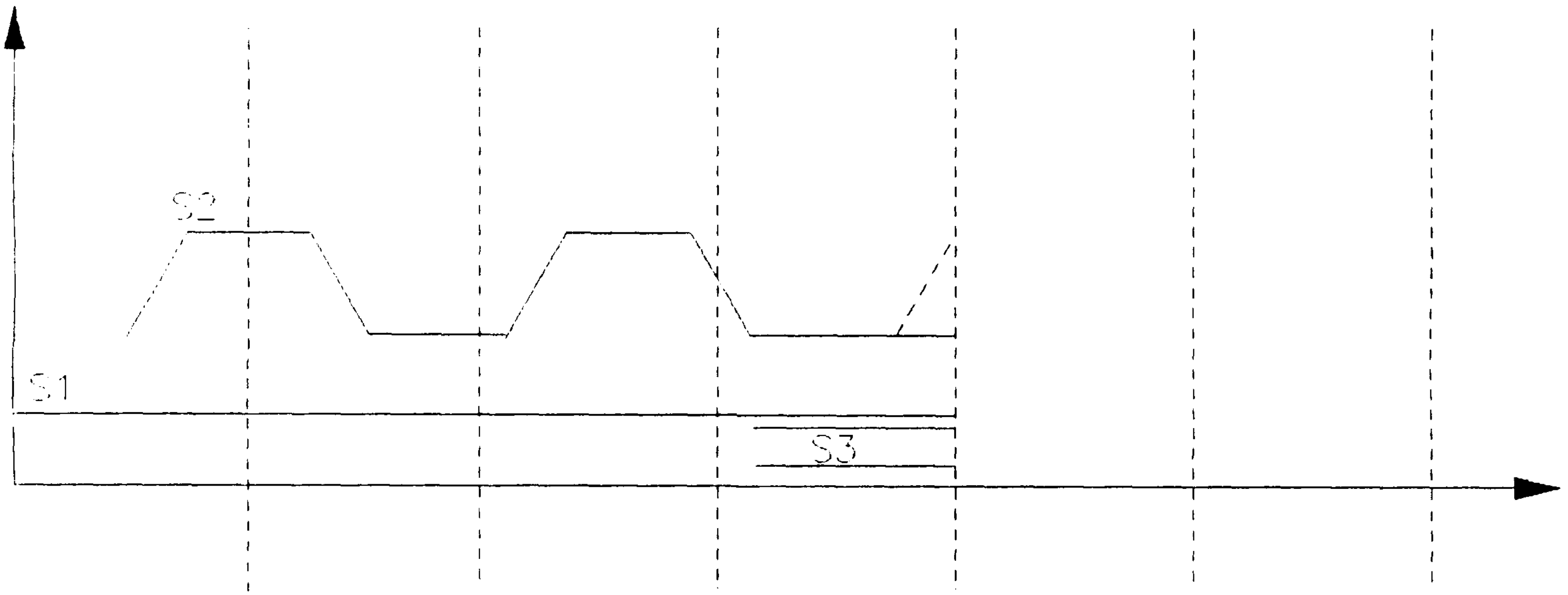


Figure 7: Front-End Signal Processing in Frame Three.



Data – Data Discrepancy T.D and STFT
Diagnosis: Frequency Resolution
Signal Re-processing with increased STFT window length
and increase # of peaks detected to all peaks.

Figure 8: Front-End Signal Processing in Frame Four.



S3: has two harmonics. It belongs to Class C Targets.

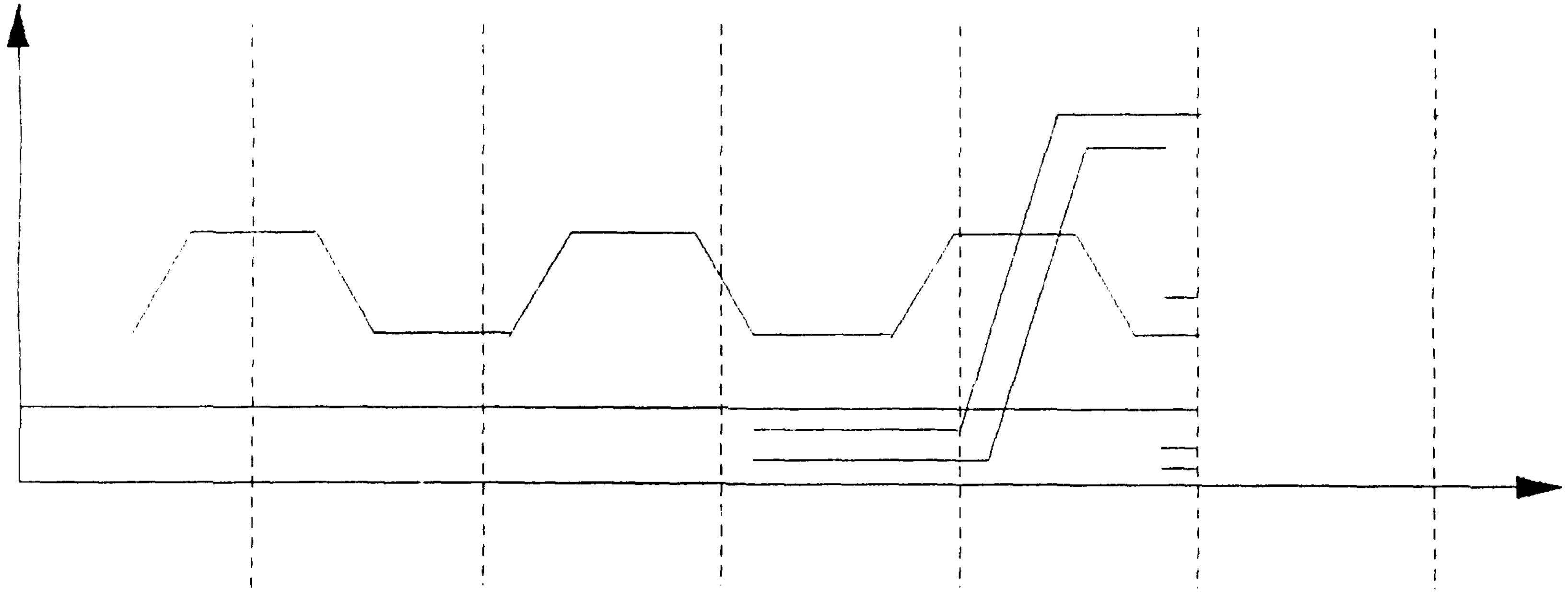
S3: Uncertainty due to incomplete temporal data and incomplete
frequency data.

Signal re-processing to prune out frequency alternatives for S3.

No support data for other frequencies. Class C pruned for S3.

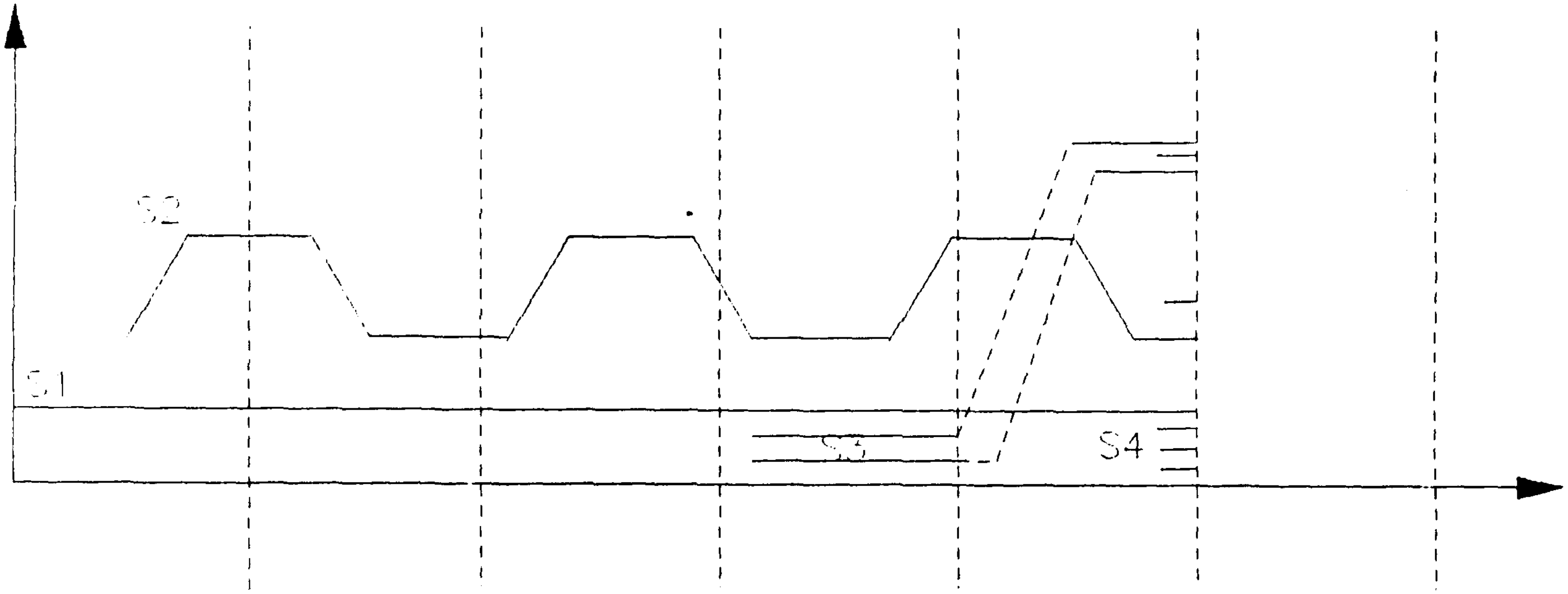
To eliminate incomplete temporal data uncertainty wait for next frame.

Figure 9: Frame Four After the First Signal Re-processing.



Global parameter adaptation to detect all peaks.
Data – Data Discrepancy TD and STFT
Diagnosis: Frequency resolution
Signal re-processing with increased STFT window length.

Figure 10: Front-End Signal Processing in Frame Five.



S3: Classified as Target D.
S4: Uncertainty because two harmonic sets (one or two sources)
S4: Uncertainty because incomplete temporal data.

Figure 11: Frame Five After the First Signal Re-processing.

6 Sound Classification Problem

To further refine the H-LASP paradigm, we picked a real-time sound classification problem which offers two major advantages: (1) it shares many low-level and high -level processing requirements with many other signal interpretation problems such as radar signal interpretation and (2) the acoustic signal database is readily available in our university laboratories for testbed experiments.

The sound classification problem for which our testbed is designed arises in the context of real-time interpretation of acoustic signals received by a system (robot, if you will) stationed in a household environment. This means that the various sounds being received by the system have to be classified in terms of the sources from which these sounds originate. In the household environment, we are interested in sources such as telephones, vacuum cleaners, babies, speech, footsteps, doorbells etc. Such sounds may be simultaneous both in time and frequency.

The goal of the sound classification system is to associate sound sources with portions of the acoustic waveform received by the system. The real-time requirement imposed on the system is that sources should be associated with portions of the waveform within a time frame that is appropriate to the goals of the overall system. For example, if the overall system is to respond to the ring of a telephone, it is necessary that the telephone ring be classified in a time frame that allows appropriate action to be taken (such as answering the telephone). Although our testbed is not designed to take such actions, it is supplied with appropriate knowledge about the time frame within which various types of sources have to be classified. There is furthermore an internal objective of an H-LASP system which also forces the classification to be done as quickly as possible: the classification of sounds is used to adapt the real-time signal processing. Finally another real-time constraint is imposed by the fact that any practical system can hold only a finite amount of data. Thus, if the sound classification is allowed to significantly lag behind the rate at which the signal information is being received, the system will be forced to lose data.

The complexity of the sound classification problem largely arises from the fact that at any given time multiple sources of sound may be present in the environment. Therefore, the signals from each of these sources overlap in time. Furthermore, in most cases there is significant overlap in their frequency content as well. The problem is further complicated because of the variability over time in the temporal and frequency characteristics of the signals received from just one source. For example, the sound of a vacuum cleaner has different characteristics depending on whether it is stationary, being pushed, or being pulled. Finally, the presence of noise in the received signals makes the problem that much harder.

To classify sounds, a system must search in both the time and frequency domai for chacteristics that help to identify particular sources and to discern between overlapping sources (such as when a telephone rings while a vacuum cleaner is being used in the background). There are many signal processing strategies that are available for transforming waveform data into various time and frequency domain representations where the search for appropriate features can be conducted. The search for these features and the construction of source hypotheses by combining such features and comparing these against knowledge about sound sources is the high-level processing component of the sound classification problem.

A practical sound classification system has a finite amount of signal processing resources. However, there is a large variety of sound sources which require their own individually tailored signal processing strategies to ensure detection of important features in the time and frequency domains. A practical sound classification system must therefore adapt its signal processing resources in accordance with its latest interpretation of the sound generating environment – a task that clearly calls for high-level adaptive signal processing.

To classify sounds, a system must possess different types of knowledge regarding sound sources. This includes knowledge about the physics of sound propagation, knowledge about the characteristics of sounds emanating from different sources (including the variability in such characteristics) and knowledge about the type of signal processing appropriate for each type of source. There is an abundance of such knowledge in the physical acoustics and psycho-acoustics literature.

7 Signal Processing for Sound Classification

In this section, we describe the signal processing resources utilized in the sound classification testbed for our project. These resources fall into three major categories: (1) Time Domain Analysis, (2) STFT analysis and (3) Filterbank Analysis.

In time-domain analysis, a time-domain waveform is analyzed for properties such as power, zero-crossing density, zero crossing spacing, and waveform envelope frequency. Estimates of the waveform power are formed by averaging the energy in the digitized samples (sampling rate in our system is 10 KHz) of the waveform over short intervals of time. The number of samples in a waveform segment used for estimating power can be varied to be as s-mall or as large as desired. Zero crossings are detected by an algorithm that searches for sign changes between consecutive waveform samples. The density is computed by calculating the number of zero-crossings in a waveform segment and dividing by the duration of the segment. The length of the segment used for this purpose is once again an adjustable parameter. For any given segment, another time domain subsytem produces the time difference between consecutive zero crossings as a function of time. From these zero-crossing spacings, the signal processing system calculates a measure of the uniformity of the zero-crossing spacings. Finally, the time-domain analysis also includes a non-linear filtering process that estimates the envelope of a waveform segment and from it calculates the frequency associated with that envelope.

In STFT (short-time Fourier transform) analysis, the system multiplies a waveform segment with a shaping window and takes the Fourier transform of the result using the FFT algorithm. Peaks in the resulting spectrum are then detected (the specific criterion used for peak detection has several adjustable parameters). Spectral peaks from consecutive (and usually partially overlapping) waveform segments are then compared. Using a decision criterion which also has several adjustable parameters, the system decides whether a peak belongs to a peak-track continuing from a previous segment or whether the peak might be the beginning of a new peak-track or whether it is just a spurious peak. Thus, the final output of the STFT analysis is in the form of peak tracks in the combined time-frequency domain.

Filterbank analysis is used in our testbed to separate the waveform into components that fall into different (although possibly overlapping) frequency bands. This allows the system to focus on frequency bands that are expected or known to have high signal-to noise ratio. It should be noted that each of the filters in the filterbank have adjustable center frequencies and bandwidths. The output of each filter is a time-domain waveform to which time-domain analysis or STFT analysis or both can be applied. The filterbank in our testbed has a total of 4 filters.

8 Blackboard Database

In interpreting source information from the acoustic waveform, it is necessary to consider certain intermediate information levels. Our initial system design requires six information levels:

- *Segment Level:* There are a variety of signal processing techniques that can be applied to the acoustic waveform to extract various types of information. In our system, we use short-time Fourier transform (STFT) analysis, time domain (TD) analysis, and filterbank (FB) analysis. These techniques are applied to waveform segments of various lengths. It is thus necessary for the sound classification system to keep track of the segments from which the higher levels of information have been extracted. This is all the more important because our system design often requires some of the waveform data to be reanalyzed in light of the higher-level information gathered with respect to that segment of the data.

- *Peak Level:* At this level, we store information about the frequency content found in the various waveform segments. This information takes the form of peaks that have frequency locations, bandwidths, power and some shape characteristics.

- *Track Level:* At this level, we represent the evolution in time of the peaks found at the lower level. Peaks found in neighboring segments are considered to belong to the same track if parameters of those peaks are close enough according to known criteria for allowable dynamics in the tracks for everyday sound sources.

- *Microstream Level:* To each acoustic source in the environment, there correspond one or more tracks. A micro-stream is a single track belonging to a particular source and is further identified in terms of three sub-regions: attack phase, steady phase, and decay phase. Each of these sub-regions have a variety of parameters associated with them in order to gain specific information about the microstream.

- *Stream Level:* . The sound from a single source typically consists of several micro-streams that are synchronized with each other. A group of synchronized micro-streams is referred to as a stream. An example of a stream would be a ring, such as that from a telephone, which typically has two dominant microstreams at two different frequencies.

- *Source Level:* At this level, sources are explicitly identified with the streams found at the lower level.

Objects at the various information levels are supported by objects at lower levels and explained by objects at higher levels. Our design of the system requires the sources of uncertainty to be explicitly associated with the supports and explanations for any of the objects. The control for the problem-solving is based on the uncertainties that the system determines to be most important to resolve at any particular time.

There are a variety of knowledge sources (KS's) for creating, verifying, and deleting hypotheses. The knowledge sources required by our system design use one or more of the following types of knowledge: signal processing, physical-acoustics, psycho-acoustics, and acoustic sources knowledge. We have not yet implemented any of the knowledge sources completely. We have instead worked with simulated KS's, with particular attention paid to their time-behavior in order to be able to use our testbed for experimentation with the real-time requirements for the processing.

Most of our implementation focus has been on the blackboard database. This section describes the implementation decisions we made with regard to: the representation of hypotheses at the various information levels, the use of links to connect related hypotheses, and the storage of information in those links regarding the uncertainty in the relationship between hypotheses.

In the GBB framework, every hypothesis is represented by a unit type. At the beginning of our project, we defined the following unit types:

- **Waveform hypothesis.** The waveform data is the input data for our system. Initially, we had one unit for every [time, power] pair in the waveform. Since that resulted in a very large number of units and since none of the signal processing algorithms required each pair to be enumerated individually we decided to view the entire waveform as just one unit.

- **Peak hypothesis.** We wanted to have five different levels of abstraction for a peak hypothesis. At first, we defined five different unit types for the peak hypothesis, all of them linked together. But later we realized that we could define just one unit for the peak hypothesis and place it in five different spaces such that each space allows access to only those parts of the hypothesis that correspond to a particular abstraction level.

- **Track hypothesis.** A track hypothesis consists of the list of peaks that comprise the track.

Each peak hypothesis is determined by applying a signal processing KS to a segment of the waveform. It was during the implementation process that we realized that to preserve the information about the correspondence between peaks and segments, we had to establish an intermediate information level between the input data and the peak level. We called it the Segment Level.

- **Segment hypothesis.** A segment represents waveform data in a time interval. Since the waveform data is going to be analyzed by three different KSs and the intervals these KSs use are not necessarily related, we defined three different kinds of segments: one for the STFT KS, one for the TD KS and one for the FB KS.

8.1 Blackboards and Spaces.

- We decided to have three different spaces in the **segment-level**, because although we have only one segment unit type, when a segment hypothesis is created, it is created for a particular type of KS. Thus, that type of knowledge source needs to search only among the units designated to its corresponding space.

- Every peak hypothesis is stored in one of five spaces. These represent the levels of abstraction for a peak hypothesis. The differences between these spaces are the dimensions. That is, the parameters we can use to retrieve a unit vary according to the space we are in.

- We have a separate **control blackboard**, because we want to have control units, which contain control plans, and we do not want those units to be stored with the data.

It should be noted that this hierarchy among the blackboards and spaces is used only because of efficiency. If all the units were stored in the same space, every time we look for a unit, we would have to search though all of them. So, it is better to keep a structure of this type.

In our application, a hypothesis can not be represented by a single unit because we do not get the final hypothesis in one step. To represent the notion of the evolution of a hypothesis, we use the concept of an extension of a hypothesis. A hypothesis has an extension when we get some new information that changes it, or simply makes it more accurate. Examples of this are:

- With a peak hypothesis. Suppose we get some information from the STFT KS. We create a peak hypothesis with this information. After a while, we get more information about that peak from the TD KS. This is not new data, it simply makes the information in the peak hypothesis more accurate. This is when we create a new extension for the peak.

- With a track hypothesis. We find that two peaks could belong to the same track (could support it) and so we create a track hypothesis. Later, we find that another peak could belong to that track, too. So we create a new extension for the track hypothesis, supported by this peak.

We decided to have two different unit types to represent a hypothesis:

- hypothesis unit type

- extension unit type

For every hypothesis we have one hypothesis unit type and as many extension unit types as needed.

We can also have multiple extensions, which represent alternative ways of interpreting the available information (with uncertainty). Furthermore hypotheses may support or explain other hypotheses but with a certain degree of uncertainty. Hypotheses that are related this way are connected by links. In our system, we wanted the links to explicitly store the sources of uncertainty associated with them. Since our version of GBB did not support links with properties, we had to implement our own links.

We thus have many units and links to represent a single hypothesis and we found that most times we are not going to use all of them. The question arose as to whether we should store all the units associated with a single hypothesis in a space. Usually, when searching through the blackboard, we only care about the last extensions of a hypothesis. Thus, only the latest extension of a hypothesis is kept in a space (making it retrievable through its parameters), while the intermediate extensions are only indirectly accessible (they are on the blackboard, linked to the latest extension, but they are not in any space).

We have found the main advantage of GBB for our application to be the flexibility it offers in making changes as the design of our system evolves. At the beginning we did not know exactly what we needed and we started with an initial blackboard structure. As we were defining the system, we found we needed to add new blackboards or spaces or that we needed to change the dimensions of a unit. With GBB, this was just a matter of changing definitions and recompiling. Here is an example of such a modification in our system:

1. At the beginning we defined only one space to store all the segment units. Later, we found that as we were going to have segments for three different KSs it could be useful to have the segments in three different spaces: one with the segments for the STFT KS, one with the segments for the TD KS and another one with the segments for the FB KS. We also wanted to add a new slot-dimension to the segment units. To make all these changes we only had to change the the segment-level space into a blackboard, define three new spaces in this blackboard. and change the segment unit definition. GBB automatically took care

of the rest (changes in retrieval functions, and so on) upon compilation.

Two difficulties we had with GBB related to links and compilation time.

1. We needed links with properties. In our version of GBB, links were just simple pointers. We therefore had to define our own links outside GBB.

2. It takes a long time to compile the definitions, particularly the unit definitions.

9 Discrepancy Detection

A major accomplishment of our project was the design of a specific discrepancy detection strategy for the sound classification testbed. Previous work on the H-LASP paradigm had largely ignored the specifics of how discrepancy detection would be accomplished in an actual system. Besides being useful for the implementation of the testbed, our design of the discrepancy detection strategy also resulted in a general framework for viewing the discrepancy detection process for any H-LASP application. In this section, we describe this general framework and illustrate it with examples from the sound classification testbed.

In the most general sense, discrepancy detection in H-LASP is concerned with comparing features of the signal processing outputs with expectations about those features based on the evolving scenario interpretation and a-priori knowledge about the application domain. In our work on the sound classification testbed, we have found that it is convenient to divide discrepancy detection into three basic categories of discrepancies: subsystem - subsystem discrepancies, subsystem-expectation discrepancies, and expectation simulation discrepancies. We describe each of these categories below.

Subsystem-subsystem discrepancies are discrepancies found between the outputs of different signal processing subsystems. For example, time-domain analysis may indicate the presence of a source at a certain frequency but the STFT analysis may not show the presence of a spectral peak track at that frequency. A number of different reasons, depending upon the parameter settings of the subsystems, may account for such a discrepancy. One possibility is that the STFT analysis may have the energy threshold (below which it ignores spectral peaks) set too high. Another reason might be that the analysis segment used by the STFT analysis is too short to allow sufficient frequency resolution to pick up the peak at that particular frequency. A third reason could be that the parameters that determine the specific criterion for associating peaks with particular tracks is not appropriate for the characteristics of the particular frequency peaks under consideration. Yet another possibility is that there really is not a source at that frequency but rather the time-domain analysis (which mostly operates under a single source assumption) gives a frequency estimate that is a hybrid produced due to the presence of sources at more than one frequency. Which particular reason applies in a specific case is determined by the diagnostic reasoning process.

Subsystem-expectation discrepancies are discrepancies found between signal processing outputs and expectations about those outputs based upon the high level scenario interpretations. For example, suppose that the system has recognized that a telephone is ringing. If a couple of rings have already taken place, the system (using its knowledge about the ringing of telephones) can predict when the next ring should take place. If the signal processing system does not produce the required features at the predicted time, this discrepancy will have to be resolved either by gathering further evidence that the telephone has stopped ringing or by checking if the signal processing parameters had not been appropriately set (this may happen if in the meantime another sound source had appeared in the environment and the signal processing resources had been refocused on that source).

Expectation-Simulation discrepancies are discrepancies between the system's expectations about what is going to happen in the signal- generating scenario at some future time and what features the signal processing outputs will have at that time (as determined by simulating the actions of the signal processing under the predicted conditions). For example, consider the situation where the system has recognized that a telephone is ringing. It might then be reasonable for the system to expect that somebody is going to answer the telephone. That would lead to an expectation that the sound of a human voice will be detected in the near future. At this point, the system can run a simulation that predicts what kinds of features the signal processing system (with its current parameter settings) would produce. If those features are not considered suitable for adequately recognizing human speech, the system may decide to readjust the signal processing parameters appropriately. It should be noted that the simulation in our testbed is carried out using the operators (that model distortions produced by the signal processing) used by the diagnosis knowledge source.

The most frequently occuring discrepancies are of the subsystem-subsystem type. An important part of designing the procedures for detecting such discrepancies is to make sure that such detection does not take place at too fine a level. Because the signal processing operations involve various degrees of approximation, a certain amount of discrepancy is always present between subsystem outputs at most given times. Although some of these discrepancies may be important to resolve, many others do not require such resolution. Since the system has to perform in real-time, it is necessary that any combinatorial explosion in the detection of dicrepancies be avoided. The discrepancy detection algorithms themselves have parameters that determine their sensitivity to various types of discrepancies. In our system, these parameters are used to constrain the number of discrepancies generated at any given time. To illustrate this idea, consider two situations involving the ringing of a telephone: in one case there is little background noise while in the other the backgroung noise is significant. In the noisy case, estimates of the loudness of the telephone as produced by the STFT and the time-domain analysis may differ considerably without there being a need to act upon that discrepancy. However, in the less noisy case, even small discrepancies may be considered a sufficient reason to explore whether or not another source has appeared. In our research so far, the issue of controlling the sensitivity parameters of the discrepancy detection has been considered in the most rudimentary ways. We feel that further research on this issue is called for in future work.

10 Diagnosis

The task of the diagnosis subsystem in the H-LASP system is to generate a simple but plausible explanation for discrepancies detected between an initial signal state and a goal signal state. The initial signal state is derived from the output of an acoustic signal processing subsystem whose output is considered to be a more accurate description of the signal environment; the goal state is derived from the output of a signal processing subsystem whose output is considered to be a less accurate representation of the signal environment due to improperly-tuned signal processing parameter settings.

The explanation is produced via a plan-and-verify strategy used in conjunction with a signal abstraction hierarchy. The abstraction hierarchy both suppresses signal information and also changes signal representation at various levels. The planning phase generates a candidate explanation by applying the generic means-ends analysis of GPS to the initial and goal states at a particular abstraction level, while the verify phase uses the entire abstraction hierarchy both to verify that the explanation satisfies the constraints of even the lowest (i.e., most complex representation) signal abstraction level and to notify the planning phase when to try applying GPS reasoning at a lower level of abstraction. An explanation takes the form of a sequence of distortion operators which maps the initial state into the goal state.

The plan-and-verify strategy begins with selecting the highest level of abstraction (i.e., the simplest representation of signal states) as the level at which to apply the GPS algorithm. This is done because by ignoring as much detail as possible, the diagnosis system can postulate explanations with as few operators as possible. In other words, the system works with simplest explanations first. The diagnosis system uses two mechanisms to prevent combinatorial explosion during the GPS search for operators to use in constructing an explanation. First, no operator is allowed to appear more than once in a particular plan. This follows from the fact that each operator represents a single process in the signal processing system; once the distortion process occurs at some point in the system, it remains in existence throughout the rest of the processing system and does not occur again.

The second mechanism for controlling GPS search is the use of an ordering relationship among classes of signal states. The classes used for the aircraft tracking application and those used for the robotic hearing application will be described later. Each operator specifies the allowable classes of input and output signal states. In an explanation, an operator cannot appear before another operator whose input signal class precedes the operator's output signal class. This considerably reduces the operator search space, but it should be noted that operators whose input and output state classes are the same can appear in any order with respect to each other.

Once an explanation has been proposed, the verify phase of the diagnostic strategy takes place. The abstraction level of the verification is the lowest one at which a description of the initial state is known. Verification proceeds as a degenerate case of the GPS algorithm at the lowest abstraction level, except that no "real" operator search is carried out: the algorithm simply selects the operators in accordance with the plan to be verified. If verification succeeds, the diagnosis system returns the explanation. If verification fails, however, the diagnosis system attempts to "patch" the explanation depending on the nature of the failure.

There are two types of explanation failures. In one, the preconditions of an operator in the explanation are not satisfied by the output state of the operator preceeding it. In this situation the diagnosis system attempts to find a sequence of operators explaining the discrepancy between the state and the preconditions of the failed operator. This patch is constructed with the GPS algorithm at the highest abstraction level which permits reasoning with the kind of signal representation at which the new discrepancy was observed. In the second type of explanation failure the output state of an operator does not match the qualitative description anticipated for it in the original explanation. In this case the failed operator is removed from the explanation and a "sub-explanation" is devised to replace its position in the explanation. In both types of failures, if no local readjustment is possible, the diagnostic system abandons the candidate explanation and starts from the planning phase again to generate a new explanation, though at a lower abstraction level than the one previously used for explanation generation.

10.1 Acoustic Localization Application

In a previous acoustic localization application [5], five abstraction levels were used: direction, power, frequency, band, and Gaussian levels. At the direction level, each signal is associated with just one characteristic–its direction in the direction spectrum. Other characteristics are hidden at this level. At the frequency level, signals are described not only in terms of direction spectra, but also in terms of their maximum and minimum frequencies. The power level represents signals in terms of their direction spectra and their net power. At the band level, power and frequency representations are combined, while the Gaussian level adds signal bandwidth information to the band level representation. Six operators were actually implemented for the diagnosis system in the aircraft tracking application, though thirteen operators had been spec-

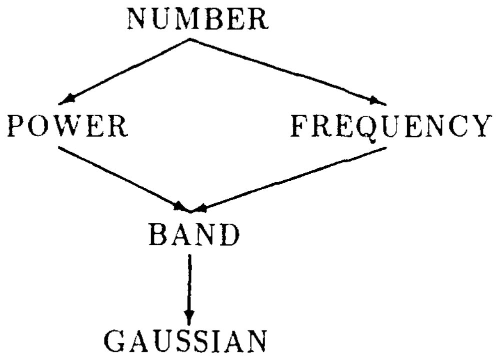


Figure 12: First Abstraction Hierarchy

ified. Consequently, the range of sophistication of explanations generated by the system was limited during system testing. The system used the following operator input/output state classifications, with their precedence order as listed: propagation, continuous-temporal, discrete-temporal, continuous-spatial, discrete-temporal, continuous-spatial, and discrete-spatial. Propagation states represent plane-wave signals propagating through the atmosphere. Continuous-temporal states represent one-dimensional analog signals, and discrete-temporal states represent one-dimensional digital signals. Continuous-spatial states represent two-dimensional analog wavenumber spectra, and finally, discrete-spatial states represent digitized wavenumber spectra.

10.2 Adapting the Diagnosis System to a New Domain

This subsection describes the changes that were made to the diagnosis system in order to apply it to the sound classification problem. Specifically, we discuss the design of a new abstraction hierarchy, the specification of new state classes, and the implementation of a new set of distortion operators.

In adapting the system to robotic hearing, we found it useful to characterize signals in terms of their prominent peaks in the frequency spectrum. An early hierarchy that was developed to support this characterization appears in Figure 1. Its levels, and their details of signal representations, were exactly

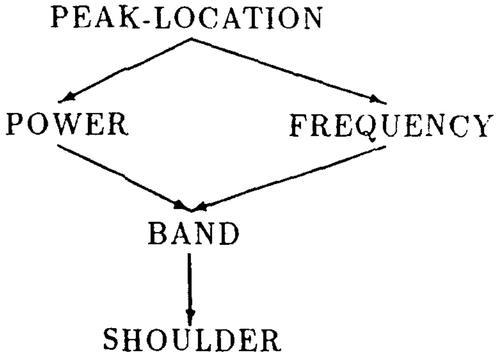


Figure 13: Second Abstraction Hierarchy for Robotic Hearing

the same as the hierarchy used in the aircraft tracking problem, except that the direction level was replaced by the number level. At this level signals were represented by the number of prominent peaks in their frequency spectra. In the course of testing the redesigned system, it was found that the number level did not support the generation of any but the most trivial explanations (e.g., only one operator). It was also found that the frequency-level representation of prominent peaks in terms of minimum and maximum frequencies was not a natural one for the problem domain. Many of the distortion operators that were specified lent themselves more naturally to characterizing peaks at the frequency level in terms of their center frequencies.

To make use of these experimental observations, a new hierarchy was developed. The names of the levels are peak-location, frequency, power, band, and shoulder. Figure 2 illustrates their refinement hierarchy. The peak-location level associates each prominent signal peak with just one characteristic: the location of the peak's center frequency in the frequency spectrum. The power level includes the power of the signal measured at the peak's center frequency along with information from the peak-location level. At the frequency level however, peaks are characterized in terms of their center frequency and their left- and right-shoulder frequencies. The band level combines the frequency and power level representations, while the shoulder level adds the measured signal powers of the frequencies at the peaks' shoulders to the band level representation.

In the sound classification domain, the signals processed by the system are not spatially-oriented in nature; they are characterized in terms of time and frequency. Hence, our diagnosis system's state classification scheme required a few adjustments. In the new scheme, the four classes used by the diagnosis system to constrain operator search are propagation, continuous-temporal, continuous-frequency, and discrete-frequency.

11 Control

The control component of high-level adaptive signal processing is required to deal with uncertainties that arise due to a number of factors. To begin with, the received signal from a source may be corrupted due to interfering signals from other sources or noise. Secondly, many of the signal processing algorithms use approximations to extract various signal features and thus introduce uncertainties. Real-time considerations sometimes force approximations in the processing and sometimes they lead to certain kinds of processing to be postponed or not to be applied at all, causing further uncertainties in the data. The higher-level processing itself has real-time limitations and thus a certain amount of focusing is inevitable in most situations. Thus, while a source that is considered important by the system may be focused upon, information about other sources may be neglected. Since a practical interpretation system retains the lower levels of data for only a finite amount of time, focusing can result in data from unclassified or partially classified sources to be lost. The consideration of such factors led us to conclude that management of uncertainty in the evidence gathered by the system has to be an important component of an H-LASP system.

For the H-LASP testbed, we have adopted a control framework [Carver] developed at the University of Massachusetts. In this framework, interpretation is modeled as a process of gathering evidence to manage uncertainty. The key components of the approach are a specialized evidential representation system and a control planner with heuristic focusing. The evidential representation scheme includes explicit, symbolic encodings of the sources of uncertainty in the evidence for the hypotheses. This knowledge is used by the control planner to identify and develop strategies for resolving the uncertainty in the interpretations. Since multiple alternative strategies may be able to satisfy goals, the control process can be seen to involve search.

Heuristic focusing is applied in parallel with the planning process in order to select the strategies to pursue and control the search. This framework allows the use of a flexible focusing scheme which can switch back and forth between strategies depending on the nature of the developing plans and changes in the domain.

The basic control loop in this framework is a goal-driven process. The highest level goal in our sound classification task is to remove uncertainties from the most recent scenario interpretation. This invokes a plan (stored in a blackboard referred to as the control blackboard) called *Remove Uncertainties from Scenario Interpretation.* In accordance with the basic control plan formalism, this control plan specifies subgoals and the order (if any) in which they are to be satisfied. In this case, there are two subgoals, to be iterated over sequentially until all sources of uncertainty are removed or the total time allocation for the process has been used. The first subgoal is to find a sound-source hypothesis on the blackboard with uncertainty in its classification. The second subgoal is to eliminate the sources of uncertainty in a specified sound-source classification hypothesis.

The first subgoal, finding a source hypothesis with uncertainty in its classification, triggers a *primitive plan.* These kinds of plans represent actions which may be carried out by a Knowledge Source (KS). In this case, a knowledge source is triggered that searches the sound-source level of the blackboard for a sound-source hypothesis that has uncertainty. The KS uses a variety of criteria to decide which hypothesis to choose. It should be noted that there will always be at least one hypothesis, namely *silence*, at the source-hypothesis level. The types of uncertainties specified along with source hypotheses include: no supporting stream-hypothesis, incomplete supporting stream-hypothesis, uncertain supporting stream-hypothesis, and alternative source hypothesis supported by same stream-hypothesis. The selected source-hypothesis is then passed over to the plan for meeting the second subgoal.

The second subgoal, to eliminate a source of uncertainty from the selected source-hypothesis, then triggers a plan called *Eliminate Sources of Uncertainty.* The control plan formalism includes the specification of input variables. In this case the input variable will take on the value of the selected hypothesis. The control plan formalism also includes output variables, whose values are bound to appropriate values and returned to the plan that called the current subplan. In our present case, there are no output variables specified, thus no values are returned after the execution of the subplan. This subplan contains two further subgoals. The first is to find the sources of uncertainty (there may be more than one) associated with the source hypothesis and the second is to eliminate a given source of uncertainty. These subgoals are iterated sequentially until all the sources of uncertainty have been dealt with. The heuristic focusing mechanism decides the order in which the sources of uncertainty are attacked if there is more than one source of uncertainty associated with a source. The sources of uncertainty are found through a knowledge source. The second subgoal has a plan consisting of several further subgoals including: to gather evidence for non-existent stream-hypothesis, to gather further data about partially-supported stream-hypothesis, to eliminate uncertainty in a stream-hypothesis, to gather evidence to resolve the conflict between multiple source hypotheses supported by the same stream-hypothesis. Which of these subgoals is pursued depends on the type of uncertainty that is to be eliminated. The selected subgoal then triggers a control plan. Sometimes, there are multiple plans available for the same subgoal. The heuristic focusing mechanism is used to decide which plan is used under the given circumstances.

The above process continues, where subgoals lead to plans and plans lead to further subgoals until primitive plans are reached. The whole process is guided by the heuristic focusing mechanism. In our case, the search for uncertainties and efforts to resolve them can reach down to the lowest levels in the blackboard, where even signal processing KS's may be triggered.

The signal processing KS's are invoked not only by the goal-driven process described above but they are also triggered by a data driven or opportunistic process that is limited in our testbed to the lowest three levels of the blackboard. Thus as the signal data arrives in the system, it triggers knowledges to create segment hypotheses, which in turn trigger knowledge sources that create peak hypotheses, and these peak hypotheses trigger knowledge sources that create track hypotheses. The hypotheses at the higher levels, microstream, stream, and sound-source, are only created by the goal-driven process.

Focusing heuristics repiesent meta-level knowledge relative to the knowledge in the control plans. Wheras control plans embody problem solving strategies for interpretation, focusing heuristics embody strategies for selecting the appropriate problem solving strategies. The focusing heuristics with particular control plans. There are several points at which focusing decisions must be made so we partition the focusing knowledge into four different classes: variable, subgoal, matching, and updating. *Variable* focusing knowledge is associated with each of the input variables of a control plan and is used to select among competing bindings for a variable. *Subgoal* focusing knowledge is used to select among multiple active subgoals for a plan instance. *Matching* focusing knowledge is used to select among the multiple plans which are applicable to satisfying a subgoal. *Updating* focusing knowledge is associated with each subgoal of a control plan and is used to decide how to proceed when a plan for satisfying the subgoal completes (i.e., succeeds or fails).

The focusing mechanism is also extended to make it possible for the system to shift its focus between competing strategies in response to the characteristics of the developing plans and factors such as data availability. Focusing is extended by allowing variable and matching focus decisions to be: absolute, postponed or preliminary. *Absolute* focusing heuristics simply select a single path to be pursued - subject of course to *potential plan failure* (which is handled by the updating process). A *postponed* focusing decision creates a refocus form which specifies the paths to be pursued, the conditions for refocusing, and a refocus handler. Refocus conditions are evaluated following the execution of any action (only actions generate new knowledge). When they are satisfied, the refocus handler is invoked and re-evaluates the choices within the new context in order to eliminate the new foci. *Preliminary* focus decisions are similar to postponed decisions except that refocusing involves a re-examination of all the original alternatives as opposed to just those that were initially focused upon. Preliminary and postponed focus decisions control the system's backtracking since they effectively define the backtrack points and the conditions under which the system backtracks.

The basic mechanisms of the control process described above have been incorporated into our testbed. We are currently implementing the specific control plans and focusing heuristics into the system.

12 Resource Allocation

The parameter adjustment component of the H-LASP paradigm may be viewed as a means for resource allocation. The need for resource allocation for the low-level processing components arises because of two factors. The first factor, the *signal variety* factor, is the enormous variety and conflicting nature of the signal processing requirements of the input signals in most signal interpretation applications, including sound classification. The signal processing resources (which are always finite) have to deal with an infinite variety of signal classes. A practical way of dealing with this problem is to parameterize the signal processing algorithms. By adjusting the parameters of an algorithm it can be made to deal with different classes of signals. The second factor that leads to the need for resource allocation is the *real-time performance* factor. In a real-time situation, there is not always enough time to do all the signal processing the system would ideally carry out. In such cases, focus-of-attention decisions have to be made about the use of the signal processing resources within the limited time frame.

The *signal variety* factor for resource allocation arises because the requirements that any particular signal type imposes on the signal processing are often in conflict with requirements of other signal categories. For example, signals whose frequency content changes rapidly as a function of time require STFT analysis whose segment length parameter is relatively small. On the other hand, signals whose frequency domain characteristics are very detailed need to have their STFT analysis done with a relatively large value for the segment length parameter. Signals that have both rapidly varying frequency characteristics as well as fine frequency domain detail would require two separate analyses; one with a short segment length and the other with a long segment length. Another example of conflicting signal processing requirements can be seen in situations that involve the presence of multiple signals. In such situations it becomes necessary to separate the contributions due to individual signals. How signals are separated from each other depends on the nature of the individual signals. Thus a signal processing system requires some information about the nature of the individual signals in order to tailor its processing for the purpose of separating signals. An alternative in this case would be not to attempt to separate the signals at the signal-processing stage, but rather to attempt separation of sound source characteristics at the higher levels of processing. A problem with such an approach is that if signal separation is not accomplished at the lower levels, the interference between signals (which is linear) usually leads to non-linear interactions between signal features at the higher levels. Such non-linear interactions are generally more difficult to resolve.

To illustrate the *real-time* factor leading to the need for resource allocation, consider a situation where signals from two sources are being received by the system. Furthermore, let us assume that the two signals have different time-frequency characteristics and thus require different parameter settings for the STFT analysis to be performed on them. If the real-time constraints force the system to perform just one STFT analysis, it is forced to choose between the two signals. This allocation of the STFT resource would have to be based upon the importance attached to the classification of the individual signals as well as previous progress made by the system in classifying the signals. If such considerations do not lead to a clear choice, an alternative is to time-slice the STFT analysis of the two signals. That is, the system goes back and forth between the two signals, focusing on the STFT analysis of each over disjoint time intervals.

13 Real-time Considerations

An important consideration in building the sound classification testbed has been to ensure that the processing strategies can be applied under real-time constraints. In this section, we discuss how the knowledge sources associated with the blackboard framework are designed to handle the real-time requirements.

In deciding this, we realized that we had five different types of KSs depending on how we could assign a processing time to them. These are:

- **FIXED TIME knowledge sources.** These are the ones that always require the same amount of time and this time is known before the KS is run.
- **MAX-TIME knowledge sources.** We do not know how long these KSs are going to take until they have finished their work. But, because of their characteristics we do know the maximum time they are going

to take. (These have to search through the database, but this is a finite search).

- **AVG-TIME knowledge sources (average-time).**

  We do not know how long these KSs are going to take until they have finished and we do not have a maximum time for them as they do a heuristic search through the database. What we do have for these KS's is an average of how long they take.

- **APPROX-WITHIN-TIME knowledge sources.** These KSs are those that have a time restriction for its execution. Since they can not spend as much time as they may need, a level of abstraction is selected depending on how much time they have. In other words, if they have very few time, they will work in a high level of abstraction because in this level they will consider less data and so the processing will be faster.

- **RESTRICTED-TIME knowledge sources.** These KSs have a time restriction for their execution, but in this case no abstraction is possible. So, these KSs will do as murh as they can in the time they have. It is possible that they will not get any useful result within their time restrictions.

14 Bibliography

1). L. Erman, F. Hayes-Roth, V. Lesser, and R. Reddy. The Hearsay II speech understanding system: integrating knowledge to resolve uncertainty. *Computing Surveys,* 12 June, 1980.

2). P. Nii, E. Feigenbaum, J. Anton, and A. Rockmore. Signal-to-symbol transformation: HASP/SIAP case study. *The AI Magazine*, pp. 23-35, Spring 1982.

3). N. Carver and V. Lesser. Planning for the Control of an Interpretation System. COINS Technical Report 89-39. University of Massachusetts. April 1989.

4). D. Corkill. GBB Reference Manual. COINS Technical Report 88-66. University of Massachusetts. July 1988.

5). S.H. Nawab, V. Lesser, and E.E. Milios, "Conceptual Diagnosis of Signal Processing Systems," IEEE Trans. Systems, Man, and Cybernetics, Special Issue on Diagnostic Reasoning, May/Jun. 87.

MISSION

of

Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control, Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. The areas of technical competence include communications, command and control, battle management information processing, surveillance sensors, intelligence data collection and handling, solid state sciences, electromagnetics, and propagation, and electronic reliability/maintainability and compatibility.